ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

VIROLOGY

Virology 307 (2003) 255–265

www.elsevier.com/locate/yviro

Proteomics computational analyses suggest that hepatitis C virus E1 and pestivirus E2 envelope glycoproteins are truncated class II fusion proteins

Robert F. Garry[a,*] and Srikanta Dash[b]

[a] *Department of Microbiology and Immunology, Tulane University Health Sciences Center, 1430 Tulane Avenue, New Orleans, LA 70112, USA*

[b] *Department of Pathology and Laboratory Medicine, Tulane University Health Sciences Center, 1430 Tulane Avenue, New Orleans, LA 70112, USA*

Received 20 September 2002; returned to author for revision 16 October 2002; accepted 19 October 2002

Abstract

Class II fusion proteins encoded by tick-borne encephalitis virus (TBEV), dengue virus, and Semliki Forest virus have a fusion peptide located at the end of a rod-like molecule comprised of three antiparallel β sheet domains. Proteomics computational analyses suggest that hepatitis C virus (HCV) envelope glycoprotein E1 and pestivirus envelope glycoprotein E2 are truncated class II fusion proteins. Similarities were also detected between the receptor-binding portion of TBEV E and HCV E2, and between TBEV small membrane protein precursor prM and pestivirus E1. The proposed models of Flaviviridae envelope proteins can facilitate drug and vaccine development.

Keywords: Viral fusion proteins; Hepatitis C virus envelope glycoproteins; Pestivirus; Proteomics; Glycoprotein structure; Virus evolution

Introduction

Entry of enveloped animal viruses requires fusion between the viral membrane and a cellular membrane, either the plasma membrane or an internal membrane. Class I fusion proteins possess a "fusion peptide" at or near the amino terminus, a pair of extended α helices and, generally, a cluster of aromatic amino acids proximal to a hydrophobic transmembrane anchoring domain (Carr and Kim, 1993; Suarez et al., 2000; Wilson et al., 1981). Several otherwise disparate viruses, including orthomyxoviruses, paramyxoviruses, retroviruses, arenaviruses, and filoviruses, encode class I fusion proteins varying in length and sequence, but highly similar in overall structure (Gallaher, 1996; Gallaher et al., 1989). X-ray crystallography of the E glycoprotein of tick-borne encephalitis virus (TBEV), a member of the genus flavivirus of the Flaviviridae family, revealed a structure for this fusion protein distinct from other fusion proteins (Rey et al., 1995). E possesses an internal fusion peptide stabilized by dicysteine linkages and three domains (I–III) comprised mostly of antiparallel β sheets. In the slightly curved rod-like configuration of the E protein present in the virion, the fusion peptide is located at the tip of domain II, the furthest point distal from the C-terminal transmembrane anchor. Examination by Lescar et al. (2001) of E1, the fusion protein of the Togavirus Semliki Forest virus (SFV), revealed a remarkable fit to the scaffold of TBEV E. Recently, E of dengue virus, a medically important flavivirus, was also shown to have a class II structure (Kuhn et al., 2002). Based on sequence similarities, it is likely that the E glycoproteins of other members of the flavivirus genus within the family Flaviviridae, including West Nile virus, are also class II fusion proteins. Proteomics computational analyses presented here suggest that glycoproteins of viruses from members of the other two genuses of the Flaviviridae family, hepaciviruses and pestiviruses, have differently truncated class II fusion protein structures.

Results

Proteomics computational analyses suggest that hepatitis C virus E1 is a truncated class II fusion protein

The Flaviviridae family consists of three genuses, flaviviruses, hepaciviruses and pestiviruses. Hepatitis C virus, the only member of the hepacivirus genus, encodes two

* Corresponding author. Fax: +1-504-584-1994.
E-mail address: rfgarry@tulane.edu (R.F. Garry).

0042-6822/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0042-6822(02)00065-X

envelope glycoproteins, E1 (gp35) and E2 (gp70), both with C-terminal transmembrane anchor domains. Yagnik et al. (2000) used proteomics computational analysis tools to produce a model of HCV E2 templated on the known structure of TBEV E. All known neutralization antibody binding epitopes could be placed at surface locations in the proposed model. Lescar et al. (2001) stated that their structural determinations of SFV E1, which established the existence of a second class of fusion proteins, "indeed support the proposed model of the hepatitis C virus envelope protein E2 which was based on the 3D structure of the flavivirus envelope protein E." However, Flint et al. (1999) provided strong evidence that the E1 glycoprotein contains the fusion peptide of HCV. Therefore, we considered the possibility that E1, in addition to or rather than E2, might have a class II fusion protein structure.

Previous studies have successfully used a "Rosetta Stone" strategy employing the fusion peptide and other identifiable features in combination with computer algorithms that predict secondary structure to construct useful working models of viral envelope proteins. Gallaher and coworkers' model of the retroviral transmembrane glycoprotein (TM) (Gallaher et al., 1989) was based on the known structure of HA2, the prototypic class I fusion protein (Wilson et al., 1981). Later, Gallaher (1996) fit the fusion protein of Ebola virus, a filovirus, to retroviral TM. Both models proved remarkably similar to the structures eventually solved by X-ray crystallography (Chan et al., 1997; Malashkevich et al., 1999; Weissenhorn et al., 1998; Weissenhorn et al., 1996).

A similar approach, supplemented with newer proteomics computational tools, was applied to HCV E1, which we modeled to fit the scaffold of TBEV E, the prototypic class II fusion protein. Because HCV E1 is shorter than TBEV E, we reasoned that the former might contain several "deletions" relative to the latter. The HCV E1 fusion peptide (Flint et al., 1999) was assumed to be located at the end of the molecule farthest from the carboxyl terminal (C-terminal) transmembrane anchor domain, and, like other class II fusion proteins, to be comprised mostly of antiparallel β sheets. This latter assumption was supported by Chou-Fasman (Chou and Fasman, 1974) and Robson-Garnier (Biou et al., 1988) analyses, the most commonly applied secondary structure prediction algorithms. (Chou-Fasman and Robson-Garnier analyses of TBEV E and HCV E1 are posted as supplemental material on All the Virology on the World Wide Web; http://www.virology.net)

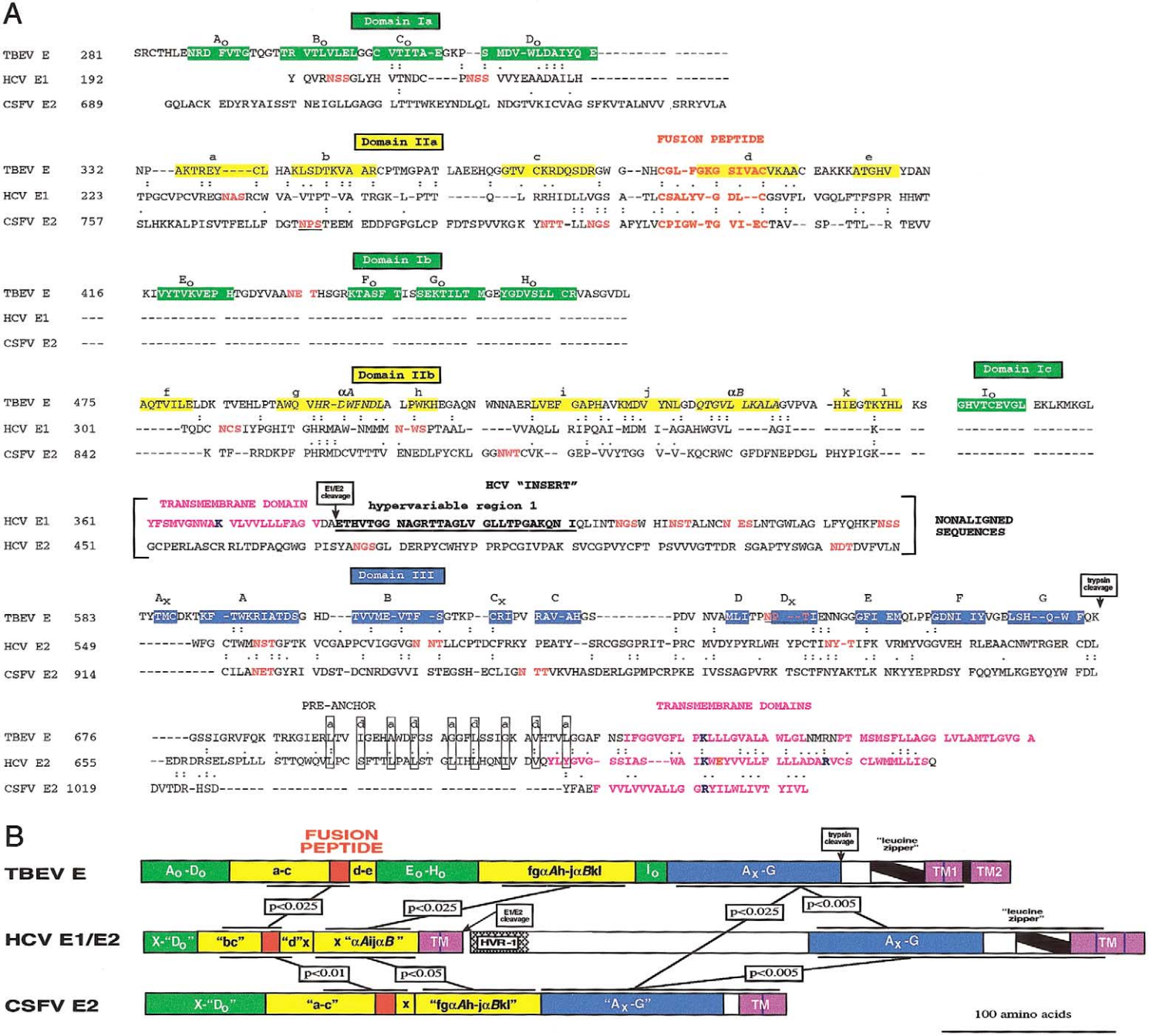
To begin the comparison, the putative fusion peptide of HCV (amino acids [aa] 272 to 281 of the full-length polyprotein) was aligned with the fusion peptide of TBEV E (aa 385–396) (Fig. 1A). Both TBEV E and HCV E1 fusion peptides have cysteine residues at either end and contain a core of mostly aromatic and hydrophobic aa (Fig. 1A, red). Another domain readily identifiable in HCV E1 is the transmembrane domain. Amino acids 361 to 381 of the hydrophobic sequence near the carboxyl terminus of E1 were predicted to form a transmembrane helix by TMpred (TMpred score 1308, >500 is statistically significant).

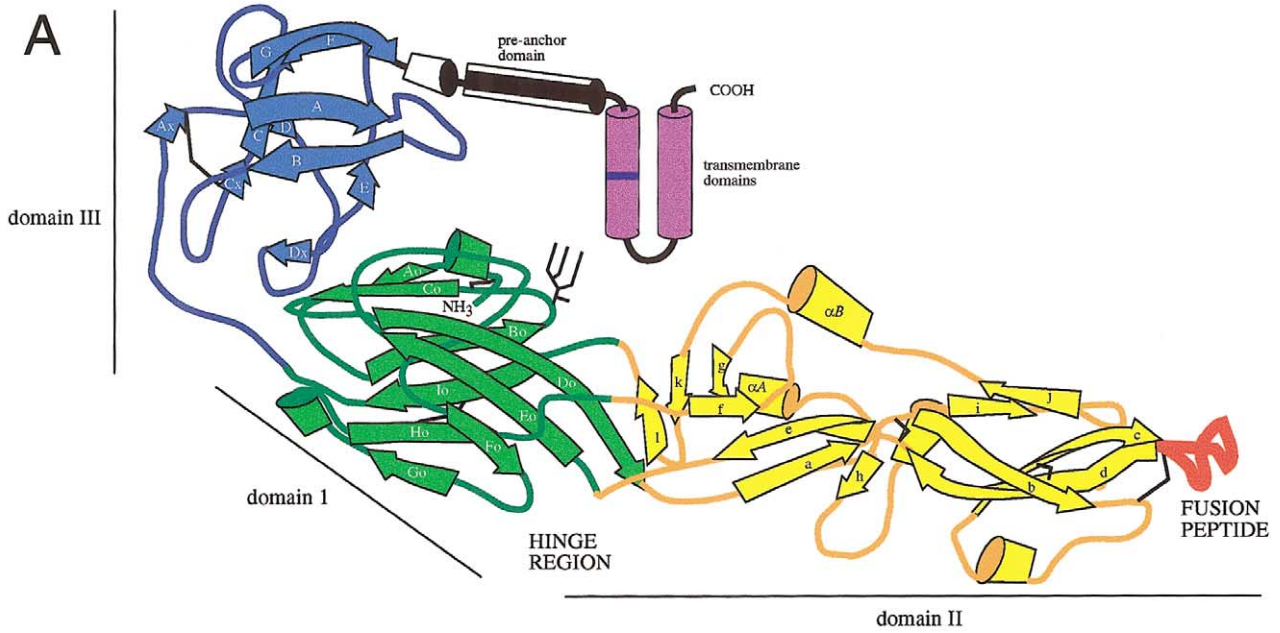
The sequence similarities between TBEV E and HCV E1 do not permit overall alignment by computational methods alone. However, several regions of predicted β sheets and α helices in HCV E1 showed similarities to sequences known to assume those secondary structures in TBEV E (Fig. 1A). Beginning from the amino terminus, the first similarity of HCV E1 begins in β sheet D_o of TBEV E and extends through the fusion peptide. PRSS3, a sequence alignment algorithm, was used to confirm that there is a significant similarity ($P < 0.025$) between aa 246–281 of HCV E1 and aa 350–396 of TBEV E (Fig. 1B). The fusion peptide is flanked by β sheets in class II fusion proteins and predicted β sheets with similarities to the b and c β sheets of TBEV E are indeed predicted to be present on either side of the putative HCV E1 fusion peptide by Chou-Fasman and Robson-Garnier analysis. HCV E1 also has an extended region of similarity with the amino acid sequence between the two longest helices in TBEV E, αA and αB. There is a statistically significant ($P < 0.025$) alignment of aa 316–356 of HCV E1 with aa 496–544 of TBEV E (Fig. 1B).

To determine the plausibility of these alignments, a three-dimensional model of HCV E1 was scaffolded on domain II of TBEV E (Fig. 2A). Similar sequences/structures were drawn in similar locations. Reorienting the "b" sheet in E1 is the only change relative to E required to bring the eight cysteine residues into close proximity. The four dicysteines of HCV E1 potentially form a "zipper" down the center of the molecule like the three dicysteines in domain II of TBEV E (Fig. 2B). This model locates the five HCV E1 glycosylation sites so they are surface accessible. Additionally, most of the hydrophobic residues are present in a region on one side of E1 between the fusion peptide and the transmembrane anchor (see below, Fig. 5). Collectively, these results suggest that HCV E1 is a truncated class II fusion protein.

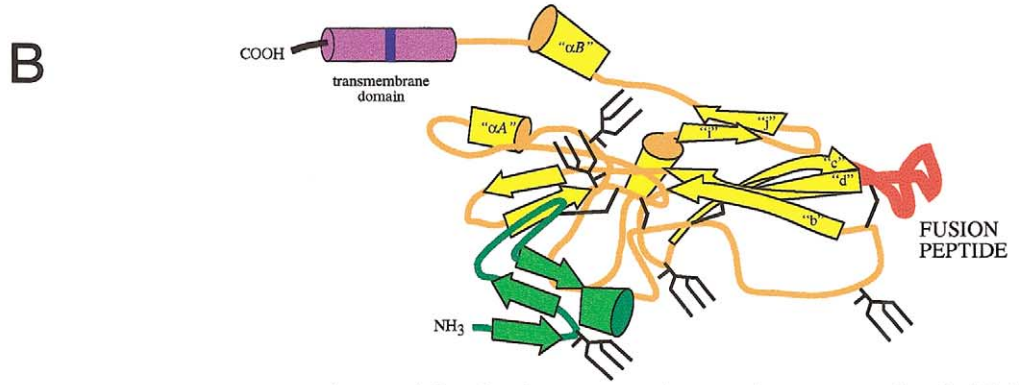
Each of the HCV E1 structures drawn in Fig. 2B conforms to both Chou-Fasman and Robson-Garnier predictions, with the exception of the region from "i" to "αB." The structures designated "i" and "j" were predicted to be β sheets by Chou-Fasman analysis, but α helical by Robson-

Fig. 1. Alignments of tick-borne encephalitis virus E, hepatitis C virus E1, and classical swine fever virus E2. Panel A: Alignments were constructed as detailed in the text. Amino acids are numbered from the beginning of the TBEV, HCV and CSFV polyproteins in this and subsequent figures. Bracketed HCV insert sequences are wrapped and do not represent an alignment comparison. (:) refers to identical amino acids. (.) refers to chemically similar amino acids. Panel B: Linear arrangement of the domain structure of TBEV E as determined by Rey et al. (1995). Regions of significant sequence similarities to TBEV E in HCV E1 and E2 and CSFV E2 as determined by the PRSS3 sequence alignment program are indicated. Probabilities (*P* values) are based on 1000 shuffles.

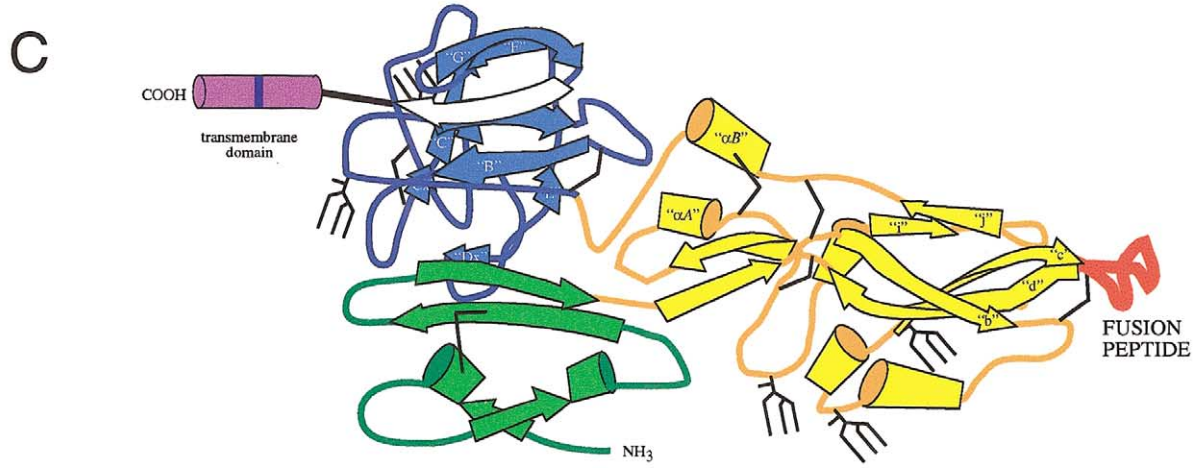

A
Domain Ia
Domain IIa
FUSION PEPTIDE
Domain Ib
Domain IIb
Domain Ic
HCV "INSERT"
TRANSMEMBRANE DOMAIN
E1/E2 cleavage
hypervariable region 1
NONALIGNED SEQUENCES
Domain III
trypsin cleavage
PRE-ANCHOR
TRANSMEMBRANE DOMAINS
TBEV E 281
HCV E1 192
CSFV E2 689
TBEV E 332
HCV E1 223
CSFV E2 757
TBEV E 416
HCV E1 ---
CSFV E2 ---
TBEV E 475
HCV E1 301
CSFV E2 842
HCV E1 361
HCV E2 451
TBEV E 583
HCV E2 549
CSFV E2 914
TBEV E 676
HCV E2 655
CSFV E2 1019
B
FUSION PEPTIDE
trypsin cleavage
"leucine zipper"
TBEV E
HCV E1/E2
CSFV E2
p<0.025
p<0.025
p<0.025
p<0.005
p<0.01
p<0.05
p<0.005
E1/E2 cleavage
HVR-1
TM
100 amino acids



tick-borne encephalitis virus envelope glycoprotein (E)



hepatitis C virus envelope glycoprotein 1 (E1)



classic swine fever virus envelope glycoprotein 2 (E2)

Garnier analysis. The structure designated "αB" was predicted to be a β sheet by Chou-Fasman analysis, but α helical by Robson-Garnier analysis. HCV E1 appears to be missing, relative to TBEV E, much of the portion of the molecule prior to the transmembrane helix (pre-anchor). This region of TBEV E follows the trypsin cleavage site at aa 395 used to generate that portion of the ectodomain of E examined by X-ray crystallography, and therefore, the TBEV E pre-anchor (stem) structure is uncertain. The pre-anchor of TBEV E has been predicted to form amphipathic α helices (Allison et al., 1999). A sequence (aa 693–721) of the pre-anchor domain in TBEV E has the characteristics of a leucine zipper, ie, leucine or other hydrophobic aa in the first and fourth (a and d) positions of a seven aa periodicity (Fig. 1A). The pre-anchor sequence of HCV E1 was also predicted to be an α helix with characteristics of a "leucine zipper" (Charloteaux et al., 2002). Because of the significant aa sequence similarity with TBEV E, the HCV E1 secondary structures between "αA" and "αB" were depicted as in TBEV E. There are several possible alternatives to the 3D model of HCV E1 drawn in Fig. 2B, and it is possible that the secondary structures change on interaction with membranes.

In contrast to HCV E1, our analyses did not reveal any sequences of HCV E2 with significant similarity to any sequence in domains I or II of TBEV E or any other flavivirus E protein (representatives of each of the four major serogroups were examined). Most of the N-terminal half of HCV E2, which include hypervariable region 1 (HVR 1), is without any sequence similarity to TBEV E. However, we detected a significant alignment ($P < 0.025$) of the C-terminal half of HCV E2 (aa 549–726) with the region of TBEV E (aa 590–763) from domain III through the first of two predicted transmembrane spanning domains of TBEV E (Fig. 1, TBEV E TM1, aa 448–469, TMpred: 1496; TM2, aa 474–496, TMpred: 1962). As discussed above, the pre-anchor region of TBEV E has a sequence (aa 693–721) with features of a leucine zipper; a similar motif (aa 675–703) is found in the HCV E2 pre-anchor (Fig. 1). in addition, the carboxyl (C) terminus of HCV E2, like that of TBEV E, contains a stretch of hydrophobic amino acids that potentially could span the membrane twice. The transmembrane anchor(s) of HCV E2 (TMpred score: 1364) is interrupted by charged amino acids like TM1 of TBEV E. Thus, by sequence alignments and structural predictions there are demonstrable similarities between the C-terminal portions of HCV E2 and TBEV E.

Proteomic computational analysis suggests that pestivirus E2 is a truncated class II fusion protein

To provide additional evidence for the HCV E1 class II fusion protein model, we determined whether the fusion proteins of the third Flaviviridae genus, pestiviruses, might share structural/sequential similarities with fusion proteins of members of the flavivirus and hepacivirus genuses. Pestiviruses encode three envelope glycoproteins, E^{rns}, E1 and E2. E^{rns}, a secreted protein with RNAse activity, does not have a hydrophobic transmembrane anchor domain. E^{rns} does possess a C-terminal charged amphipathic segment that can mediate translocation of E^{rns} across bilayer membranes (Langedijk, 2002). Pestivirus E1 and E2 both have C-terminal hydrophobic domains that could function as transmembrane anchors. Therefore, we postulated that either pestivirus E1 or E2 must be the pestivirus fusion protein.

A putative fusion peptide (aa 818–828) is present in CSFV E2, containing a consensus sequence with aromatic and hydrophobic aa located between two cysteine residues (Fig. 1). The cysteine residues as well as the sequences in between are highly conserved among pestiviruses, as is true of fusion peptides from other enveloped RNA viruses of class I and II (not shown). Although statistically significant alignments were not detected between the N-terminus of CSFV E2 and TBEV E (or between other flaviviruses), a significant alignment ($P < 0.01$) was detected between CSFV E2 (aa 792–835) and HCV E1 (aa 253–294) in this region (Fig. 1B). Furthermore, sequences flanking the putative fusion peptide were predicted to form β sheets by both Chou-Fasman and Robson-Garnier analyses (supplemental data). A significant alignment ($P < 0.05$) between CSFV E2 (aa 841–913) and HCV E1 (aa 301–383) was also determined. By extension, the central portion of CSFV E2 is predicted to structurally resemble domain II of TBEV E. A significant alignment ($P < 0.005$) was detected between aa 914–1018 of CSFV E2 and a sequence in domain III of TBEV E (aa 587–685) (Fig. 1B). There was also a significant similarity ($P < 0.005$) of this region of CSFV E2 (aa 914–1123) with a sequence (aa 549–743) in the region of HCV E2 that aligns with TBEV domain III. In addition, TMpred confirmed that the hydrophobic C-terminal domain of CSFV E2 has a high propensity to span the lipid bilayer (score: 1137). Like the transmembrane domains of HCV E1/E2 and TBEV TM1, the putative transmembrane anchor of CSFV E2 has a central positive charge.

On the basis of the regions of significant sequence sim-

Fig. 2. Models of hepacivirus E1 and pestivirus E2 based of the structure of tick-borne encephalitis virus E. Panel A. Structure of TBEV E as determined by Rey et al. (1995) is shown schematically (traced from a RasMac rendering). Panel B: A model fitting HCV E1 to the structure of TBEV E. HCV E1 sequences with similarity to TBEV E sequences are enclosed in quotation marks. Panel C: A model fitting CSFV E2 to the structure of TBEV E. TBEV E domains I–III are colored green, yellow/orange, blue, respectively, with the fusion peptide red. Similar structures are color-coded the same way in HCV E1 and CSFV E2. Black lines: dicysteine linkages. Plum amino acids or black stick figures: glycosylation sites (underlined site with central proline often not be used). Black and white structures were predicted as described in the text. A black stripe indicates a "leucine zipper." Violet: transmembrane domains. Blue amino acids: positively charged.

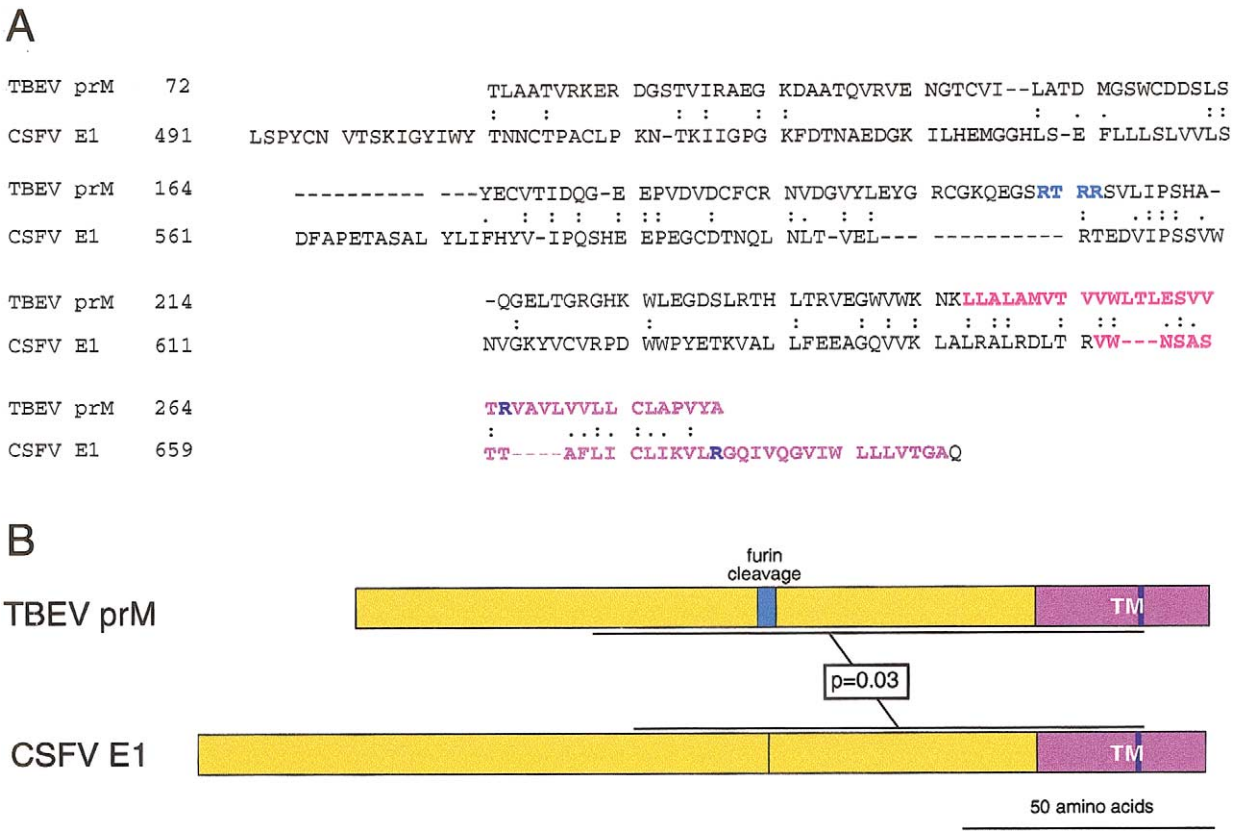


Fig. 3. Alignments of the precursor of tick-borne encephalitis virus small membrane protein, prM, and classical swine fever virus E1. Panel A: alignments were constructed as detailed in the text. Panel B: Linear arrangement of TBEV prM and CSFV E1 with a region of sequence similarity determined by the PPSS3 algorithm indicted.

ilarities between CSFV E2, HCV E1/E2 and TBEV E, coupled with the internal location of a possible fusion peptide, we conclude that relative to TBEV E, CSFV E2 is lacking a portion of domain I including segments corresponding to β sheets E_o through I_o. CSFV E2 also appears to contain a somewhat shorter segment relative to TBEV E in the pre-anchor domain, ie, the sequence between the alignment with TBEV E domain III and the transmembrane domain (Fig. 1B). No leucine zipper is evident in the pre-anchor of CSFV E2. A three dimensional model of CSFV E2 (Fig. 2C) confirms that the alignment in Fig. 1 is plausible. Each of the cysteine residues is in proximity to other cysteine residues and potentially form dicysteine bridges. Like HCV E1, CSFV E2 conforms to the structure of a truncated class II fusion protein, albeit with fewer truncations relative to flavivirus E than HCV E1. Because E1 is conserved among the pestivirus genus, the similarities of CSFV E2 with TBEV E extend to other pestiviruses.

None of the E1 envelope glycoproteins of any pestivirus bear any significant sequence similarities to any sequenced flavivirus E protein. Immature flavivirus virions contain a precursor prM to the small membrane protein M. prM is cleaved in the endoplasmic reticulum by furin or by a furin-like protease during virus release to produce the mature M protein localized on the surface of flavivirus virions (Stadler et al., 1997). A sequence (aa 173–256) of CSFV E1 has similarity ($P = 0.030$) to aa 583–654 of TBEV prM (Fig. 3A). CSFV E1 does not contain the sequence RXR/KR, the furin consensus cleavage site. CSFV E1 also does not contain an identifiable fusion peptide, although TMpred predicts a significant transmembrane spanning domain in the first third of CSFV E1. Like the transmembrane domains of TBEV E, HCV E1 and E2 and CSFV E2, and TBEV prM (TMpred score = 1828), the C-terminus of CSFV E1 is predicted to form a membrane spanning domain (TMpred score = 1884) with a central positive charge.

Gene order of Flaviviridae genomes

Genes that encode proteins with similar functions may be present in similar locations in genomes of different members of the Flaviviridae family. The positive-polarity single-stranded RNA genomes of all members of the Flaviviridae are translated into a single large polyprotein that is subsequently cleaved by viral and cellular proteases into functional proteins. The order (from N to C terminus) of proteins in the polyproteins of TBEV and other members of the flavivirus genus is C-prM-E-nonstructurals (C: capsid), and the order of proteins in the polyproteins of hepaciviruses is C-E1-E2-p7-nonstructurals (Fig. 4). The 5′ portion of the flavivirus E gene encodes the fusion peptide in domain II of

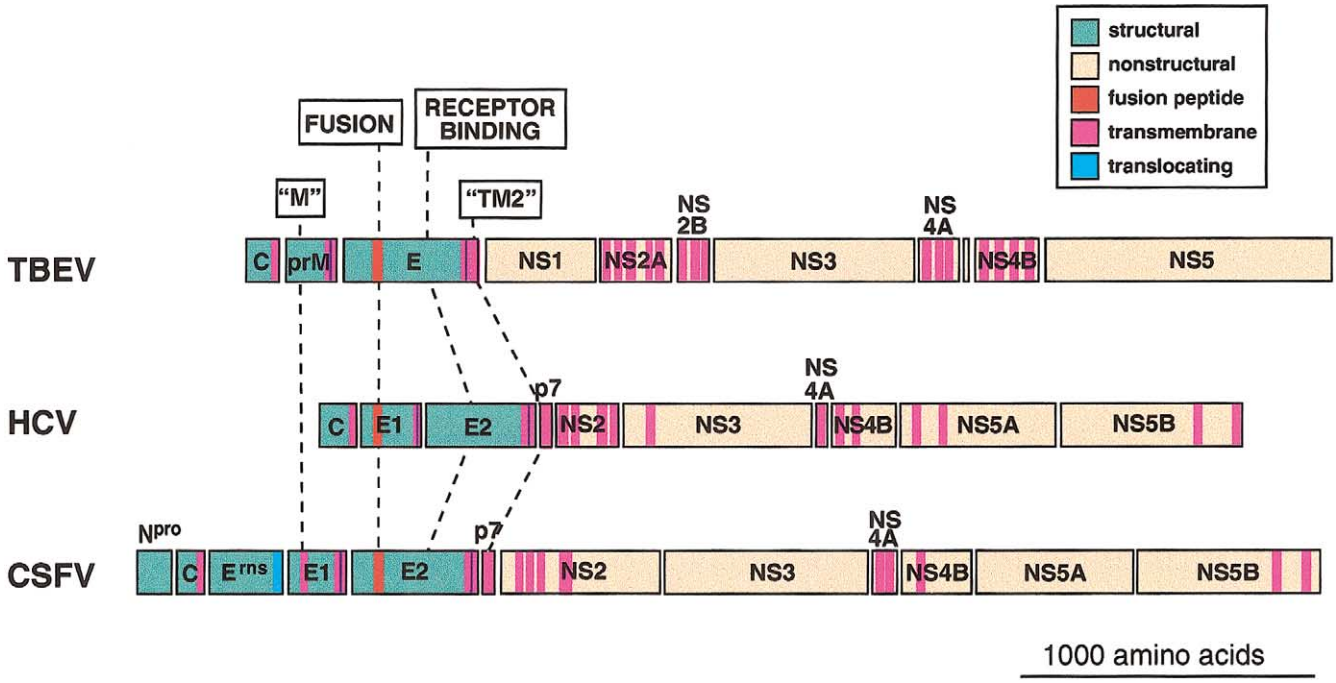


Fig. 4. Common order of proteins in Flaviviridae polyproteins. Proteins or portions of proteins with similar functions are located in similar locations along the polyproteins of members of the Flaviviridae. Hydrophobic domains were predicted using TMpred.

the E protein, whereas the receptor binding domain of E is probably located in domain III encoded by the 3′ portion of the E gene (Crill and Roehrig, 2001; Mandl et al., 2000). Fusion and receptor functions may reside in two different HCV proteins, E1 and E2 respectively, occurring in the same order as the domains of flavivirus E that carry out these functions (Fig. 4). Hepacivirus E1 and E2 may have arisen by insertion of a transmembrane anchor and variable domains, including hypervariable region 1 (HVR-1, Fig. 1), into the ancestral E gene. Alternatively, HCV E1 could have evolved into a separate fusion protein from an ancestral prM, with concurrent loss of the fusion peptide and fusion functions in E2. The sequence similarities between TBEV E and HCV E1 and E2, however, do not favor this latter possibility.

The order of proteins in pestivirus polyproteins is Npro-C-Erns-E1-E2-p7-nonstructurals. Pestiviruses encode two proteins, Npro and E^{rns}, with no obvious homologs among members of the other two Flaviviridae genuses. Pestivirus E1 and E2 are similar in sequence to flavivirus M and E, respectively. Like TBEV E, pestivirus E2 may serve both as fusion protein and receptor binding protein. These functions are carried out by TBEV E domains II and III that appear to be represented by similar structures in pestivirus E2 (Fig. 4). TBEV PrM/M functions to protect internal cellular membranes from fusion mediated by E2, and it is possible that pestivirus E1 serves the same function for E2, the fusion/receptor protein. Excepting Npro and E^{rns}, the order of structural proteins with sequence and other similarities is analogous in pestiviruses and flavivirus polyproteins.

TBEV E has two hydrophobic C-terminal transmembrane domains, TM1 and TM2 (Fig. 1). Hepaciviruses and pestiviruses encode a small hydrophobic peptide, p7, which could associate with cellular or viral membranes. The cleavage that produces p7 is inefficient and delayed, and therefore much of HCV E2 and pestivirus E2 are present in the cell as uncleaved E2-p7 precursors (Harada et al., 2000). The p7 gene is located in a similar genomic location and could have evolved from the sequence encoding the second transmembrane domain, TM2, of flavivirus E (Fig. 4). The consensus Flaviviridae genome can therefore be represented as X1-C-X2-M-fusion-binding-TM1-TM2-nonstructurals-3′, where X 1 and X2 represents inserted sequences in pestiviruses, N^{pro} and E^{rns}, respectively, M represents flavivirus prM/M-pestivirus E1, and TM2 is the second transmembrane domain of flaviviruses and p7 of hepaciviruses and pestiviruses. These similarities in gene order and functions support the hypothesis that HCV E1 is the fusion protein of HCV.

Membrane interfacial domains in a class I fusion protein and HCV E1

Although the overall structures of class I and II fusion proteins are distinct, they may share structural/functional characteristics in the parts of the molecules that interact with and disrupt bilayer membranes. It is well established that class I fusion proteins have a fusion peptide at the amino terminus of the molecule that is critical for fusion (Gallaher, 1987, 1996; Gallaher et al., 1989, 2001). Class II fusion proteins have an internal fusion peptide that are located after secondary structural folding at distal locations from the transmembrane anchor (Kuhn et al., 2002; Lescar

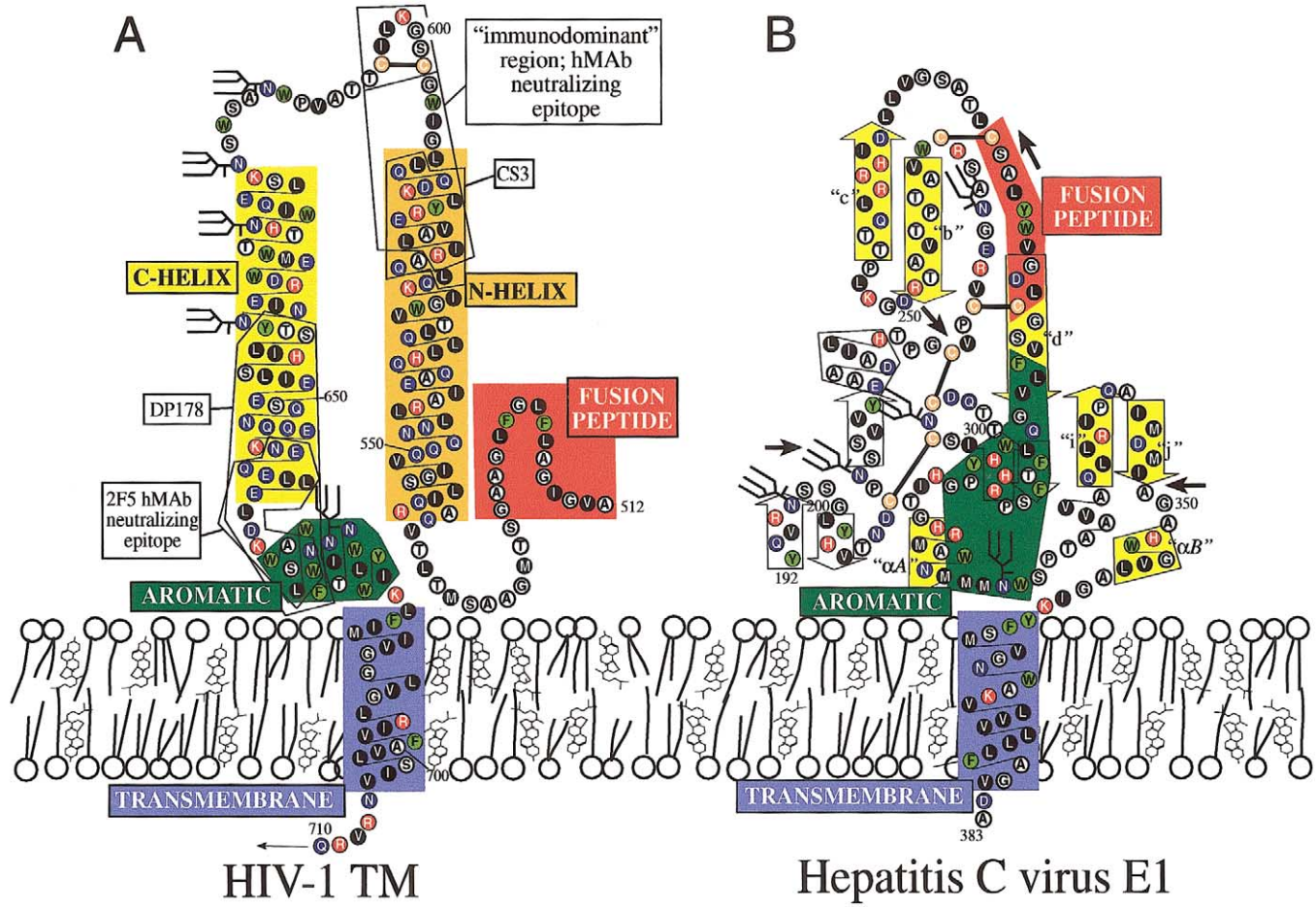


Fig. 5. Comparison of human immunodeficiency virus transmembrane glycoprotein (TM with hepatitis C virus envelope glycoprotein 1 (E1). Panel A: an updated model of HIV-1 TM from Gallaher et al. (1989) with structural motifs indicated in rainbow order. Amino acids are numbered from the beginning of the Env polyprotein. HIV-1 TM is truncated after the transmembrane domain. The precise ends of the TM N- and C-helices are unclear because of conflicting structural data. No attempt was made to align the N- and C-helices, although points of contact are solved in the coiled-coil formations. Positions of known neutralizing epitopes on TM are indicated, as well as sequences corresponding to peptides CS3 and DP178 (T20) (Qureshi et al., 1990; Wild et al., 1994) that inhibit HIV-1 infectivity. Panel B: model of HCV E1 with motifs that are shared with HIV-1 TM. Boxed arrows are predicted beta sheet structures; those colored yellow are similar to the indicated β sheets of TBEV E. Predicted alpha helical structures are outlined. Arrows denote directions that the HCV E1 structure could fold in three dimensions.

et al., 2001; Rey et al., 1995). To provide further support for the proposed models of HCV E1 and pestivirus E2, we used another proteomics computational tool to compare other potential membrane interactive domains in the proteins with the HIV-1 transmembrane glycoprotein (TM), a class I fusion protein. Besides fusion peptides, another motif in class I fusion proteins that can be important in virus:cell fusion is an aromatic as rich motif proximal to the anchor (Fig. 5A, aa 667–683, green) (Suarez et al., 2000). The pre-anchor domains of class I fusion proteins are not highly hydrophobic according to the Kyte-Doolittle hydropathy prediction algorithm; however, these domains have a tendency to partition into bilayer membranes, as revealed by analyses using the Wimley-White interfacial hydrophobicity scale (Suarez et al., 2000; Wimley and White, 1996). HCV E1 contains three domains that produce significant Wimley-White partition scores using Membrane Protein eXplorer (Jaysinghe et al., 2000). One of these is the transmembrane anchor (aa 361–372). The other two sequences with significant Wimley-White partition scores are located immediately following the fusion peptide (aa 284–300) and at a location (aa 321–340) that the model in Figure 2B predicts to be near the bilayer membrane (Fig. 5B, green). These two HCV E1 domains, in conjunction with the fusion peptide and the transmembrane anchor, potentially form a continuous track of membrane interactive regions that could channel the movement of lipids during virion:cell fusion. These Wimley-White partition analyses thus provide additional support for the proposal that E1 is the fusion protein of HCV.

Discussion

The use of proteomics computational tools, including aa sequence alignments and algorithms that predict protein

structure/function, suggests that the ectodomain of HCV E1 is a truncated version of the class II fusion protein structure first determined by Rey et al. (1995) for TBEV E. The ectodomain of HCV E1 is roughly equivalent to the part of TBEV E from the "hinge" region to the fusion peptide (Fig. 2). E2 envelope glycoproteins of pestiviruses also exhibited local sequence similarities and other predicted structures similar to TBEV E and HCV E1, allowing the determination that pestivirus E2 proteins are also likely to be truncated class II fusion proteins, although with fewer truncations relative to TBEV E than HCV E1. Significant alignments of E1 of hepatitis GB virus (GBV-B) with HCV E1 suggest that this unclassified member of the Flaviviridae family also encodes a truncated class II fusion protein (not shown).

Our conclusions contrast with those of Yagnik et al. (2000), who predicted that HCV E2 fits the scaffold of a complete class II fusion protein. HCV E1 contains a fusion peptide (Flint et al., 1999), a required motif in both class I and II fusion proteins, but HCV E2 has no amino acid sequence that is likely to function as a fusion peptide. The sequence predicted by Yagnik et al. (2000) to occupy the end of E2 most distal from the transmembrane domain, which is the location of other class II fusion peptides, does not contain a fusion peptide motif. Class I fusion proteins are generally more highly conserved among strains of a given virus than the corresponding receptor protein. The most variable domain of the two HCV envelope proteins, HVR-1, is located in E2 (Fig. 1) (Weiner et al., 1991). The placement of the cysteine residues also fails to support the model proposed by Yagnik and co-workers because the cysteines are spaced such that it would be difficult to form disulfide linkages. Our results do suggest that there are sequence and structural similarities between HCV E2 and TBEV E. However, these similarities are limited to the C-terminal portions of these proteins, and our alignments are different that those proposed previously (Yagnik et al., 2000).

Hepaciviruses, like alphaviruses, appear to use one envelope protein for attachment (E2) and another for fusion (E1). In contrast, E glycoproteins of TBEV, dengue virus, and other members of the flavivirus genus mediate both receptor binding and membrane fusion functions. E2 functions as one of the pestivirus receptor-binding proteins (Hulst and Moormann, 1997), and if the current analysis is correct, also carries out the virion:cell fusion function. In addition to E, flaviviruses encode a membrane protein prM whose functions may include shielding of cellular membranes from the fusion peptide of E (Kuhn et al., 2002). Functions of the flavivirus small membrane protein may be vested in E1 of pestiviruses, which has significant sequence similarity with flavivirus prM. Mature flavivirus virions contain prM that has been cleaved to M. Unlike M, pestivirus E1 does not associate with the virion envelope as a precursor protein and lacks a furin cleavage site.

HCV E1 may represent the minimal class II fusion protein structure required to mediate virion:cell fusion. Cell entry of flaviviruses and togaviruses that encode class II fusion proteins occurs via the endocytic route, where exposure to low pH may trigger several conformational changes in the envelope proteins and in the virion itself. In the case of TBEV and dengue virus, E converts from a dimer to a trimer bending upward to expose the fusion peptide. The β sheets of E may also rearrange to form a closed β barrel that can insert into the vesicular membrane in the manner of porins, a family of bacterial toxins. Similar changes probably occur during SFV entry, except that exposure of the fusion peptide on E1 may depend on a pH-induced conformational change in envelope protein, E2 (Lescar et al., 2001). A variation on this model posits that the pH change permits bending of the class II fusion proteins at the flexible "hinge" region between domains I and II, elevating the fusion peptide so that it can insert into the host membrane (Fig. 2A) (Lescar et al., 2001). Class II fusion proteins may even adapt a "snap-back" configuration, as proposed for class I fusion proteins, in which the fusion peptide moves closer to domain III, bringing the virus and cell membranes into contact and providing energy to drive the fusion reaction. HCV E1 and perhaps pestivirus E2 may not require these latter rearrangements. Instead, energy for driving membrane fusion may be derived from a shift from an open to a closed β barrel. Energy for the fusion reaction may also be derived from increasing favorable protein:lipid interactions, or from multimerization of the proteins into a ring surrounding the fusion pore as occurs with the upright spike glycoprotein (G) of rhabdoviruses (Carneiro et al., 2002; Roche and Gaudin, 2002).

The models proposed here are supported by the observations that envelope glycoproteins with significant sequence similarities (HCV E1/2, TBEV E and pestivirus E2, and TBEV prM and pestivirus E1) are in analogous locations in the polyproteins encoded by the three genuses of the Flaviviridae. These results suggest that members of the Flavivirus family may have a common ancestor. Divergence of the genes for the fusion proteins within the three genuses of this family may have occurred either through acquisition of sequences and/or lose of sequences in a cassette manner constrained by the domain organization of class II fusion proteins.

While the proteomic computational tools employed here have a good predictive value, X-ray crystallography will be required to determine whether HCV E1/E2 or pestivirus E2 have any structural similarities to class II fusion proteins. Development of a cell culture system that produces useful quantities of HCV particles is a prerequisite to defining the arrangement of the envelope glycoproteins on the virion surface as has been done for flaviviruses and togaviruses (Kuhn et al., 2002; Lescar et al., 2001). Diffractable crystals of HCV E1 or E2 and highly productive systems for HCV propagation are not currently available. In this regard, better drugs to treat HCV infection and an effective vaccine to prevent HCV infection are urgently needed. In the United States alone an estimated 4 million people are already in-

fected by HCV, four times the number infected by HIV-1. HCV is an important cause of liver failure and hepatocellular carcinoma. A substantial portion of HCV-infected individuals show little or no response to treatment with interferons and/or ribovirin, the only approved therapeutics. Models such as proposed here can provide useful hypotheses to guide experimental strategies for development of vaccines or drugs to prevent or treat HCV infection. Prior to the availability of X-ray structural data (Wild et al., 1993; Wild et al., 1994), several potent HIV-1 TM inhibitors were developed based on the Gallaher HIV-1 TM fusion protein model (Gallaher et al., 1989). DP178 (T20) peptide (Fig. 5A) has been shown to substantially reduce HIV-1 load in AIDS patients in preliminary results from phase III clinical trials (Hoffman-La Roche and Trimeris, 2002). Analogous peptides may be designed to block HCV virion:cell fusion. Drug and vaccine strategies targeted to the HCV envelope proteins may be broadly applicable to other members of the Flaviviridae.

Materials and methods

Sequences

Prototype strains of representatives of the three Flaviviridae genuses were used for sequence and structural comparisons. The strains examined include TBEV strain Neudoerfl (accession number: P14336), human prototype strain H (subtype 1a) of hepatitis C virus (P27958), and the Alfort 187 strain of classical swine fever virus, also known as hog cholera virus (CAA61161). Some comparisons used representatives of the major serogroups of flaviviruses, including Japanese encephalitis virus, strain JaOARS982 (P32886), yellow fever virus, strain 17D-204 (PI9901), dengue virus type 2, strain PR-159/S1 (P12823), and West Nile virus, strain NY 2000-crow3356 (AF404756). Type species of other pestiviruses, including bovine viral diarrhea virus (BVDV) genotype 1 aka pestivirus type 1, stain NADL (CAB91847), and border disease virus strain BD31 (AAB37578), were used in other comparisons. We also compared HCV sequences to those of GB virus–B virus (AAC54059), an unassigned member of the Flaviviridae.

Proteomics computational methods

Methods to derive general models of surface glycoproteins have been described previously (Gallaher et al., 1989). MacMolly (Soft Gene GmbH, Berlin) was used to locate areas of limited sequence similarity and to perform Chou-Fasman and Robson-Garnier analyses. PRSS3, a program derived from rdf2 (Pearson and Lipman, 1988), which uses the Smith-Waterman sequence alignment algorithm (Smith and Waterman, 1981), was used to determine the significance of protein alignments. PRSS3 is part of the FASTA package of sequence analysis programs available by anonymous ftp from ftp.virginia.edu. Default settings for PRSS3 were used, including the blosum50 scoring matrix, gap opening penalty of 12, and gap extension penalty of 2. The alignments presented are those that produced the highest alignment scores, rather than the longest sequences that produced significant scores. Chou-Fasman and Robson-Garnier algorithms predict protein structures in an aqueous environment, but they cannot predict protein structures in a lipid bilayer. Domains with significant propensity to form transmembrane helices were identified with TMpred (ExPASy, Swiss Institute of Bioinformatics). TMpred is based on a statistical analysis of TMbase, a database of naturally occurring transmembrane glycoproteins (Hofmann and Stoffel, 1993). Sequences with a propensity to partition into the lipid bilayer were identified with Membrane Protein eXplorer from the Stephen White laboratory (Jaysinghe et al., 2000) using default settings. RasMac, developed by Roger Sayle, was used to render 3D models of TBEV E.

Acknowledgments

This research was supported by NIH Grants AI 34764, CA54576 and CA89121.

References

Allison, S.L., Stiasny, K., Stadler, K., Mandl, C.W., Heinz, F.X., 1999. Mapping of functional elements in the stem-anchor region of tick-borne encephalitis virus envelope protein E. J. Virol. 73, 5605–5612.

Biou, V., Gibrat, J.F., Levin, J.M., Robson, B., Garnier, J., 1988. Secondary structure prediction: combination of three different methods. Protein Eng. 2, 185–191.

Carneiro, F.A., Bianconi, M.L., Weissmuller, G., Stauffer, F., Da Poian, A.T., 2002. Membrane recognition by vesicular stomatitis virus involves enthalpy-driven protein-lipid interactions. J. Virol. 76, 3756–3764.

Carr, C.M., Kim, P.S., 1993. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. Cell 73, 823–832.

Chan, D.C., Fass, D., Berger, J.M., Kim, P.S., 1997. Core structure of gp41 from the HIV envelope glycoprotein. Cell 89, 263–273.

Charloteaux, B., Lins, L., Moereels, H., Brasseur, R., 2002. Analysis of the C-terminal membrane anchor domains of hepatitis C virus glycoproteins E1 and E2: toward a topological model. J. Virol. 76, 1944–1958.

Chou, P.Y., Fasman, G.D., 1974. Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. Biochemistry 13, 211–222.

Crill, W.D., Roehrig, J.T., 2001. Monoclonal antibodies that bind to domain III of dengue virus E glycoprotein are the most efficient blockers of virus adsorption to Vero cells. J. Virol. 75, 7769–7773.

Flint, M., Thomas, J.M., Maidens, C.M., Shotton, C., Levy, S., Barclay, W.S., McKeating, J.A., 1999. Functional analysis of cell surface-expressed hepatitis C virus E2 glycoprotein. J. Virol. 73, 6782–6790.

Gallaher, W.R., 1987. Detection of a fusion peptide sequence in the transmembrane protein of human immunodeficiency virus. Cell 50, 327–328.

Gallaher, W.R., 1996. Similar structural models of the transmembrane glycoproteins of Ebola and avian sarcoma viruses. Cell 85, 1–2.

Gallaher, W.R., Ball, J.M., Garry, R.F., Griffin, M.C., Montelaro, R.C., 1989. A general model for the transmembrane proteins of HIV and other retroviruses. AIDS Res. Hum. Retro. 5, 431–440.

Gallaher, W.R., DiSimone, C., Buchmeier, M.J., 2001. The viral transmembrane superfamily: possible divergence of Arenavirus and Filovirus glycoproteins from a common RNA virus ancestor. BMC Microbiol. 1, 1.

Harada, T., Tautz, N., Thiel, H.J., 2000. E2-p7 region of the bovine viral diarrhea virus polyprotein: processing and functional studies. J. Virol. 74, 9498–9506.

Hoffman-La Roche, Trimeris, 2002. Roche and Trimeris announce 24-week results from second pivotal study of HIV fusion inhibitor T-20. http://www.trimeris.com/news/pr/2002/020516.html.

Hofmann, K. Stoffel, 1993. TMbase—a database of membrane-spanning segments. Biol. Chem. Hoppe-Seyler 374, 166.

Hulst, M.M., Moormann, R.J., 1997. Inhibition of pestivirus infection in cell culture by envelope proteins E(rns) and E2 of classical swine fever virus: E(rns) and E2 interact with different receptors. J. Gen. Virol. 78, 2779–2787.

Jaysinghe, S., Hristova, K., White, S.H., 2000. Membrane Protein Explorer. http://blanco.biomol.uci.edu/mpex.

Kuhn, R.J., Zhang, W., Rossmann, M.G., Pletnev, S.V., Corver, J., Lenches, E., Jones, C.T., Mukhopadhyay, S., Chipman, P.R., Strauss, E.G., Baker, T.S., Strauss, J.H., 2002. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell 108, 717–725.

Langeduk, J.P., 2002. Translocation activity of C-terminal domain of pestivirus Erns and ribotoxin L3 loop. J. Biol. Chem. 277, 5308–5314.

Lescar, J., Roussel, A., Wien, M.W., Navaza, J., Fuller, S.D., Wengler, G., Rey, F.A., 2001. The fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. Cell 105, 137–148.

Malashkevich, V.N., Schneider, B.J., McNally, M.L., Milhollen, M.A., Pang, J.X., Kim, P.S., 1999. Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-A resolution. Proc. Natl. Acad. Sci. USA 96, 2662–2667.

Mandl, C.W., Allison, S.L., Holzmann, H., Meixner, T., Heinz, F.X., 2000. Attenuation of tick-borne encephalitis virus by structure-based site-specific mutagenesis of a putative flavivirus receptor binding site. J. Virol. 74, 9601–9609.

Pearson, W.R., Lipman, D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444–2448.

Qureshi, N., Coy, D., Garry, R., Henderson, L.A., 1990. Characterization of a putative cellular receptor for HIV-1 transmembrane glycoprotein using synthetic peptides. AIDS 4, 553–558.

Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. Nature 375, 291–298.

Roche, S., Gaudin, Y., 2002. Characterization of the equilibrium between the native and fusion-inactive conformation of rabies virus glycoprotein indicates that the fusion complex is made of several trimers. Virology 297, 128–135.

Smith, T.F., Waterman, M.S., 1981. Identification of common molecular subsequences. J. Mol. Biol. 147, 195–197.

Stadler, K., Allison, S.L., Schalich, J., Heinz, F.X., 1981. Proteolytic activation of tick-borne encephalitis virus by furin. J. Virol. 71, 8475–8481.

Suarez, T., Gallaher, W.R., Agirre, A., Goni, F.M., Nieva, J.L., 2000. Membrane interface-interacting sequences within the ectodomain of the human immunodeficiency virus type 1 envelope glycoprotein: putative role during viral fusion. J. Virol. 74, 8038–8047.

Weiner, A.J., Brauer, M.J., Rosenblatt, J., Richman, K.H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q.L., Houghton, M., Han, J.H., 1991. Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. Virology 180, 842–848.

Weissenhorn, W., Carfi, A., Lee, K.H., Skehel, J.J., Wiley, D.C., 1985. Crystal structure of the Ebola virus membrane fusion subunit, GP2, from the envelope glycoprotein ectodomain. Mol. Cell 2, 605–616.

Weissenhorn, W., Wharton, S.A., Calder, L.J., Earl, P.L., Moss, B., Aliprandis, E., Skehel, J.J., Wiley, D.C., 1996. The ectodomain of HIV-1 env subunit gp41 forms a soluble, alpha-helical, rod-like oligomer in the absence of gp120 and the N-terminal fusion peptide. EMBO J. 15, 1507–1514.

Wild, C., Greenwell, T., Matthews, T., 1993. A synthetic peptide from HIV-1 gp41 is a potent inhibitor of virus-mediated cell-cell fusion. AIDS Res. Hum. Retro. 9, 1051–1053.

Wild, C.T., Shugars, D.C., Greenwell, T.K., McDanal, C.B., Matthews, T.J., 1994. Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. Proc. Natl. Acad. Sci. USA 91, 9770–9774.

Wilson, I.A., Skehel, J.J., Wiley, D.C., 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289, 366–373.

Wimley, W.C., White, S.H., 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat. Struct. Biol. 3, 842–848.

Yagnik, A.T., Lahm, A., Meola, A., Roccasecca, R.M., Ercole, B.B., Nicosia, A., Tramontano, A., 2000. A model for the hepatitis C virus envelopeglycoprotein E2. Proteins 40, 355–366.